

## A NEW WEB-BASED ENGINE FOR SUMMARIZING PRODUCT REVIEWS AND CUSTOMER OPINIONS

Mrs .J. SHOBANA

Assistant professor, SRM University, Chennai

**ABSTRACT** - *The current research is focusing on the area of Opinion Mining also called as sentiment analysis due to sheer volume of estimation rich web resources such as discussion forums, review sites and blogs are available in digital form. It is also known as opinion mining, mood extraction, sentiment analysis ,appraisal mining and emotion analysis. It is the important aspect for capturing public opinion about product preferences, marketing campaigns, political movements, social events furthermore company strategies. One important problem in sentiment analysis of product reviews is to produce summary of opinions based on product features. I have surveyed and analyzed in this paper, various sentiment classifications and techniques that have been developed for the key tasks of opinion mining. And I also summarized the issues and challenges of opinion mining that change the results of opinion mining.*

### 1. INTRODUCTION

The internet provides its users with immense quantities of functional information. This makes retrieval of data from different locations a strenuous job. Artificial intelligence in automated systems can be used to anatomize, synopsise and classify the extracted data. This process helps in the decision making process of various enterprises and individuals. Opinion Mining is one type of NLP, known as Natural Language Processing which involves the keeping a track of user's temperament, opinions, sentiments and emotions.

In today's virtual world rather than only providing reviews and comments on existing information users share their own ideas and thoughts on social networks and micro blogging websites. These result in the generation of huge volumes of data which in turn is available for every other user out there.

In recent years, we have witnessed that opinionated postings in social media have helped reshape businesses, and sway public sentiments and emotions, which have profoundly impacted on our Social and political systems. Such postings have also mobilized masses for political changes such as those happened in some Arab countries in 2011. It

has thus become a necessity to collect furthermore study opinions on the Web. Of course, opinionated documents not only exist on the Web, many organizations also have their internal data, e.g., client feedback collected from emails and call centers or results from surveys conducted by the organizations.

Opinion mining can be useful in several ways. For example, in marketing, it tracks and judges the success rate of an ad campaign or launch of new product, determine popularity of products and services with its versions also tell us about demographics which like or dislike particular features. For example, a review might be about a digital camera might be broadly positive, but be specifically negative about how heavy it is. The vendor gets a much clearer picture of public opinion than surveys or focus groups, if this kind of information is identified in a systematic way.

The technique to detect and extract subjective information in text documents is opinion mining and sentiment analysis. In general, the overall contextual polarity or sentiment of a writer about some aspect can be determined using sentiment analysis. The main challenge in this area is the sentiment classification in which the sentiment may be a judgment, mood or evaluation of an object namely film, book, product, etc which can be in the form of document or sentence or feature that can be labeled as positive or negative.

Classifying entire documents according to the opinions towards certain objects is called as sentiment classification. One form of opinion mining in product reviews is also to produce feature-based summary. To produce a summary on the features, product features are first identified, and positive and negative opinions on them are aggregated. Features are product attributes, components and other aspects of the product. The effective opinion summary, grouping feature expressions which are domain synonyms is critical. It is very time consuming and tedious for human users to group typically hundreds of feature expressions that can be discovered from text for an

opinion mining application into feature categories. Some automated assistance is needed. Opinion summarization does not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as the classic text summarization. [2]

The paper is organized into the following sections: the data sources used for opinion mining, introduces the sentiment classification, concepts of opinion mining model, various opinion mining approaches and opinion mining techniques. The last section summarizes the issues and challenges of opinion mining that affect the results of opinion mining.

## 2. DATA SOURCE

People and companies across disciplines exploit the rich and unique source of data for varied purposes. The major criterion for the improvement of the quality services rendered and enhancement of deliverables are the user opinions. Blogs, review sites and micro blogs provide a good understanding of the reception level of products and services.

### 2.1 Blogs

The name associated to universe of all the blog sites is called blogosphere. People write about the topics they want to share with others on a blog. Blogging is a happening thing because of its ease and simplicity of creating blog posts, its free form and unedited nature. We find a large number of posts on virtually every topic of interest on blogosphere. Sources of opinion in many of the studies related to sentiment analysis, blogs are used. [3]

### 2.2. Review Sites

Opinions are the decision makes for any user in making a purchase. The user generated reviews for products and services are largely available on internet. The sentiment classification uses reviewer's data collected from the websites like [www.gsmarena.com](http://www.gsmarena.com) (mobile reviews), [www.amazon.com](http://www.amazon.com) (product reviews), [www.CNETdownload.com](http://www.CNETdownload.com) (product reviews), which hosts millions of product reviews by consumers. [1]

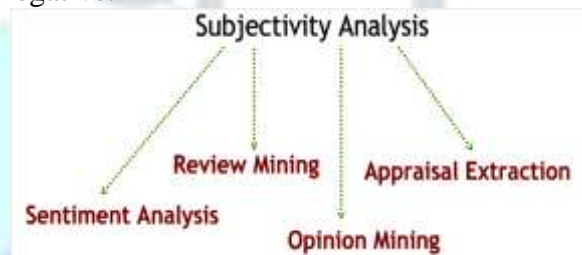
### 2.3. Micro-blogging

A very popular communication tool among Internet users is micro-blogging. Millions of messages appear daily in popular web-sites for micro-blogging such as Twitter, Facebook etc. Twitter messages sometimes express opinions which are used as data source for classifying sentiment. [4]

## 3. OPINION MINING

It is kind of web content mining. Figure 3 shows this categorization clearly.

□ **DEFINITION:** If a set of text documents (T) are given, that have opinions on an object, opinion mining intends to identify attributes of the object on which opinion have been given, in each of the document  $t \in T$  and to find orientation of the comments i.e. whether the comments are positive or negative.



Figure

4: Synonyms of Opinion mining

Figure 4 shows different terms that used interchangeably for opinion mining [5]

### 3.1 Scientific Fundamentals

#### 3.1.1 Model of Opinion Mining

As people are free to give their opinions on anything, e.g., they buy a product and then they express their views on products' features in various forums. The term „object“ is used for the entity on which comments have been given.

□ **Definition (object):** An object A is an entity. It is related to a pair,  $A: (C, R)$ , where C is the components and sub-components of A, and R is the attributes of A. Each component can have its own sub-components and attributes.

“Features” can refer to either components or attributes. It is also commonly used for objects. Let us consider a document t, which contains opinions on an object A. Generally, t is composed of sentences  $t = (s_1, s_2, s_3 \dots s_n)$ .



□ *Definition (opinion passage on a feature):* Opinion on a particular feature  $f$  of an object  $A$ , extracted from a document  $t$ , is a group of sentences in  $t$  that contain some opinion on  $f$ .

A single sentence may express opinions on several features of a product, e.g., "The picture quality of this camera is good, but the battery life is short".

□ *Definition (opinion holder):* The person giving his/her opinion on something is the holder of the opinion.

□ *Definition (semantic orientation/sentiment classification of an opinion):* The semantic orientation of an opinion on a feature  $f$  states whether the opinion is positive, negative or neutral. This classification can be done at sentence level i.e. whether a sentence contains a positive opinion on a feature of an object or it may contain negative opinion on it.

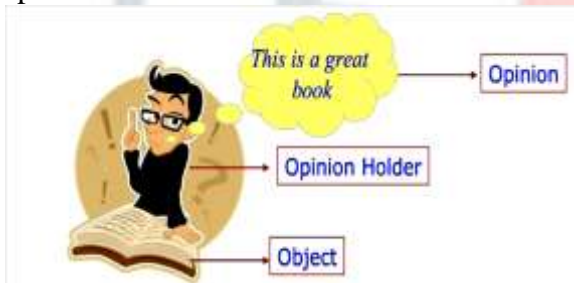


Figure 5: Model of Opinion mining

Putting things together, a model for an object and a set of opinions on the features of the object can be defined, which is called the feature-based opinion mining model[5].

## 4. SENTIMENT CLASSIFICATION

Sentiment classification mainly consists of classifying the polarity of a given text at the document, sentence or feature/aspect level expressing the opinion as positive, negative or neutral. The sentiment analysis can be performed at one of the three levels: the document level, sentence level, feature level.

*Document Level Sentiment Classification* : In document level sentiment analysis main challenge is to extract informative text for inferring sentiment of the whole document. The learning methods can be confused because of objective statements are rendered by subjective statements and complicate

further for document categorization task with conflicting sentiment. [6]

*Sentence Level Sentiment Classification:* The sentiment classification is a fine-grained level than document level sentiment classification in which polarity of the sentence can be given by three categories as positive, negative and neutral. The challenge faced by sentence level sentiment classification is the identification features indicating whether sentences are on-topic which is kind of co-reference problem [6]

*Feature Level Sentiment Classification:* Product features are defined as product attributes or components. Analysis of such features for identifying sentiment of the document is called as feature based sentiment analysis. In this approach positive or negative opinion is identified from the already extracted features. It is a fine grained analysis model among all other models [6]

## 5. OPINION MINING APPROACHES

### 5.1. Machine Learning Approaches

In general, sentiment analysis is concerned with analyzing direction based text, determining whether a text is objective or subjective and whether a subjective text contains positive or negative sentiments is a common two-class problem that involves classifying sentiments as positive or negative. Additional variations include classifying sentiments as opinionated/subjective or factual/objective. Some studies have attempted to classify emotions (such as happiness, sadness, anger, or horror) instead of sentiments. The machine-learning approach [7], treats the sentiment-classification problem as a topic-based text classification problem. Any text classification algorithm can be employed such as Naïve Bayes or support vector machines (SVMs).

### 5.2. Lexicon Based Approach

The Lexicon based approach performs classification based on positive and negative sentiment words and phrases contained in each evaluation text and mining the data requires no prior training. Two types of techniques have been used in previous semantic orientation approach based sentiment classification research: corpus-based and dictionary-based.

### 5.2.1. Corpus-based Approach

The corpus-based approach aims to find occurrence patterns of words to determine their sentiments. Researchers have proposed different strategies to determine sentiments; for example, PeterTurney [8], calculated a phrase's semantic orientation to be the mutual information between the phrase and the word "excellent" (as the positive polarity) minus the mutual information between the phrase and the word "poor" (as the negative polarity). Ellen Riloff and Janyce Wiebe [9], used a boot strapping process to learn linguistically rich patterns of subjective expressions to distinguish subjective expressions from objective expressions.

### 5.2.2. Dictionary-based Approach

Use synonyms, antonyms, and hierarchies in WordNet (or other lexicons with sentiment information) to determine word sentiments [10]. Building upon WordNet, SentiWordNet is a lexical resource for sentiment analysis that has more sentiment-related features. It assigns to each synset of WordNet three sentiment scores regarding positivity, negativity, and objectivity, respectively. SentiWordNet has been used as the lexicon in recent sentiment classification studies. The corpus-based techniques, however, often rely on a large corpus to calculate the statistical information needed to decide the sentiment orientation for each word or phrase. Therefore, they might not be as efficient as the dictionary-based techniques. Still, a good lexicon is critical for the dictionary-based techniques [5].

## 6. TECHNIQUES USED IN OPINION MINING

Database contains the important hidden information used for decision making. Different databases like relational, object oriented, transactional and spatial databases consist on the complex dataset. Major data mining techniques used to extract the knowledge and information are: generalization, classification, clustering, association rule mining, data visualization, neural networks, fuzzy logic, Bayesian networks, and genetic algorithm, decision tree, and multi agent systems, CRISP-DM model, churn prediction, Case Based Reasoning and many more.



Figure 2: Opinion Mining Technique

Rapid growth in databases has created the need to develop such technologies to extract the suggest of knowledge and information intelligently. Data mining techniques are most suitable for this purpose, these techniques directly refers Artificial Intelligence.

### 6.1. Supervised Machine Learning

Classification is most frequently used and popular data mining technique. Classification used to predict the possible outcome from given data set on the basis of defined set of attributes and a given predictive attributes. The given dataset is called training dataset consist on independent variables (dataset related properties) and a dependent attribute (predicted attribute). A training dataset created model test on test corpora contains the same attributes but no predicted attribute. Accuracy of model checked that how accurate it is to make prediction. Classification is a supervised learning used to find the relationship among attributes.

### 6.2. Unsupervised Learning

In contrast of supervised learning, unsupervised learning has no explicit targeted output associated with input. Class label for any instance is unknown so unsupervised learning is about to learn by observation instead of learn by example. Clustering is a technique used in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters.

### 6.3. Case Based Reasoning

Case based reasoning is an emerging Artificial Intelligence supervised technique used to find the solution of a new problem on the basis of past similar problems. CBR is a powerful tool of computer reasoning and solve the problems in such



a way which is closest to real time scenario. It is a recent problem solving technique in which knowledge is personified as past cases in library and it does not depend on classical rules. The past problem's solutions are stored in CBR repository called Knowledge base or Case base. Instead of solving the new problem by "first principal" reasoning, CBR use the knowledge base to reuse the solution of past similar problem if needed to the In case base repository as a new solution instance in CBR cycle consists of four R's. Nowadays it is the most emerging technique used in opinion mining systems. Statistical methods are combined with knowledge extracting techniques to enhance case searching, browsing and Reuse it for the problem solving methods semantic analysis of a sentence in natural language that can be easily used and manipulated in a textual data mining process. This sentence analysis uses and depends on several types of knowledge that are: a lexicon, a case base and hierarchy of index. In this methodology a case based reasoning model is adopted that is based on the classification rules and course of similarity for the assurance of the compliance.

## 7. OPINION MINING AND SUMMARIZATION PROCESS

Opinion Mining also called sentiment analysis is a process of finding user's opinion towards a topic. Opinion mining concludes whether user's view is positive, negative, or neutral about product, topic, event etc. Opinion mining involves analyzing user's opinion, attitude, and emotion towards particular topic. This consists of first categories text into subjective and objective information, and then finding polarity in subjective text. Opinion mining can be performed word, sentence or document level. Opinion mining and summarization process involve three main steps, first is Opinion Retrieval, Opinion Classification and Opinion Summarization.

Summarization of opinions is a major part in opinion mining process. Summary of reviews provided should be based on features or subtopics that are mentioned in reviews. Therefore, feature extraction and opinion summarization are key issues. Many researchers worked on summarization product reviews [11]. The opinion summarization process mainly involve following two approaches. One is Feature based summarization another one is Term Frequency based summarization.

## REFERENCES

- [1] G.Vinodhini and RM.Chandrasekaran, Sentiment Analysis and Opinion Mining: A Survey, Volume2, Issue 6, June 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering
- [2] Zhongwu Zhai, Bing Liu, Hua Xu and Hua Xu, Clustering Product Features for Opinion Mining, WSDM'11, February 9–12, 2011, Hong Kong, China. Copyright 2011 ACM 978-1-4503-0493-1/11/02...\$10.00
- [3] Singh and Vivek Kumar, A clustering and opinion mining approach to socio-political analysis of the blogosphere, Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference.
- [4] Alexander Pak and Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining,
- [5] Ganesan, K. A., and H. D. Kim. 2008. Opinion Mining - A Short Tutorial (Talk) , University of Illinois at Urbana Champaign.
- [6] M. Mehdi Owrang O. 2007. Discovering Quality Knowledge from Relational Databases, American University, USA.
- [7] Jebaseel A., E. Kirubakaran, "M-Learning Sentiment Analysis with Data Mining Techniques", International Journal of Computer Science and Telecommunications, Vol. 3, 2012
- [8] Turney P., "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 417-424, 2002.
- [9] Riloff E. , J. Wiebe, "Learning Extraction Patterns for Subjective Expressions", in: proceeding of Empirical Methods on Natural Language Processing Conference , pg. 105-112, 2003
- [10] Hamouda A., M. Rohaim, " Reviews Classification Using SentiWordNet Lexicon" , The Online Journal on Computer Science and Information Technology, Vol. 2, 2011
- [11] C. Chen, F. Ibekwe-SanJuan, E. SanJuan, and C. Weaver.
- [12] Visual analysis of conflicting opinions. In IEEE Symposium on Visual Analytics Science and Technology, page.